

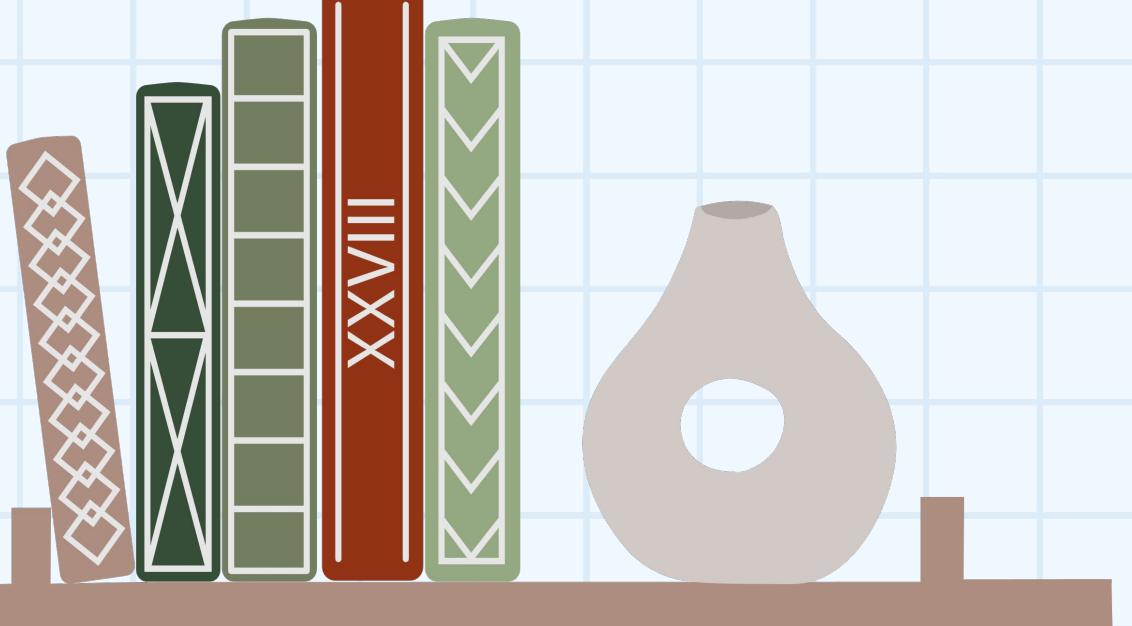
Powered by
Futurense



BS./BSC.IN

Applied ai and Data Science

Basics of Data Analytics





Powered by

Futurense

Let's dive into and learn:



1

Data Formatting



Data Formatting

- After Data cleaning, we ensure that the data has been formatted correctly
- Data Formatting includes
 - Data type conversion - Ensure values in a field are stored as data types of that field.
 - Standardizing data – Units of measurement are standard across datasets.
 - Syntax Errors – Typos, White spaces

Data Formatting



Powered by

Futurense

- While formatting data, make sure the data is formatted ***consistently***.



Data Formatting

Transaction ID	Customer ID	Product ID	Purchase Date
53666	34987	453	01-02-2025
54333	35671	984	45669.00
56473	39823	382	29-01-2025
56234	36005	226	45662.00

Change into the same format

Transaction ID	Customer ID	Product ID	Purchase Date
53666	34987	453	01-02-2025
54333	35671	984	12-01-2025
56473	39823	382	29-01-2025
56234	36005	226	05-01-2025



Data Formatting

Transaction ID	Customer ID	Product ID	Purchase Date
53666	34987	453	01 February 2025
54333	35671	984	12-01-2025
56473	39823	382	29-01-2025
56234	36005	226	05-01-2025

Change into the same format

Transaction ID	Customer ID	Product ID	Purchase Date
53666	34987	453	01-02-2025
54333	35671	984	12-01-2025
56473	39823	382	29-01-2025
56234	36005	226	05-01-2025



Powered by



Futurense

Data Formatting in

The screenshot shows the Microsoft Excel interface with the Home tab selected in the ribbon. The main area displays a blank worksheet with columns A through H and rows 1 through 14. In the top-left corner of cell A1, there is some faint, illegible text. The ribbon also includes File, Insert, Page Layout, Formulas, Data, Review, View, Help, Acrobat, and Tell me what you want. On the far right of the ribbon, there is a search bar with the placeholder text "Tell me what you want". The Number Format dropdown menu is open on the far right of the ribbon, listing various formats: General, Number, Currency, Accounting, Short Date, Long Date, Time, Percentage, Fraction, and More Number Formats... The "General" option is currently selected.

Data Formatting



Powered by

Futurense

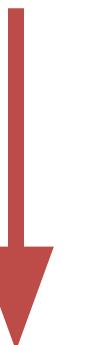
- Data formatting also requires ensuring that all the entries are in the ***same unit of measurement.***

Data Formatting



Product ID	Product Name	Price per KG
453	Banana	Rs.14
984	Potato	\$0.50
382	Watermelon	Rs.16
226	Grapes	\$ 0.45

Convert into the same unit of measurement



Product ID	Product Name	Price per KG
453	Banana	14
984	Potato	22
382	Watermelon	16
226	Grapes	29

Data Formatting



Powered by

Futurense

- Check the source data for unit of measurement if they aren't mentioned.
- Use the common-sense test!



Data Formatting

ID	Name	Age	Height
GH89765	Priya	29	63
RF23876	Rahul	32	61
UT75097	Devesh	28	162
FC37694	Bhavna	31	59



Inspect the data carefully when units of measurement may not be available.

ID	Name	Age	Height
GH89765	Priya	29	63
RF23876	Rahul	32	61
UT75097	Devesh	28	63
FC37694	Bhavna	31	59

Data Formatting



Powered by

Futurense

- Look for inconsistencies in the formats.
- Typographical errors
- Blank spaces



Data Formatting

ID	Name	Age	Height	State
GH89765	Priya	29	63	Uttar Pradesh
RF23876	Rahul	32	61	UP
UT75097	Devesh	28	63	Rajasthan
FC37694	Bhavna	31	59	Rajasthan
TD65298	Karan	41	65	Rajasthan



ID	Name	Age	Height	State
GH89765	Priya	29	63	Uttar Pradesh
RF23876	Rahul	32	61	Uttar Pradesh
UT75097	Devesh	28	63	Rajasthan
FC37694	Bhavna	31	59	Rajasthan
TD65298	Karan	41	65	Rajasthan

Recap



Powered by
Futurense

Common cases of data formatting
issues.



Powered by



Futurense

Thank you

